DOCUMENT RESUME

ED 205 590 TM 810 483

AUTHOR	Yeh, Jennie P.
TITLE	A Reanalysis of Test Use Data. Test Use Project.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C.
REPORT NO	CSE-R-143
PUB DATE	Aug 80
NOTE	59p.
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Educational Assessment; Elementary Education; *Elementary School Teachers; School Demography; *Teacher Attitudes; Teacher Education; Teacher Made Tests; *Testing; Test Results; *Test Use

ABSTRACT

The general goal of the Center for the Study of Evaluation (CSE) Test Use Project is to examine and describe the features and applications of tests and other assessment methods that contribute to the improvement of instruction. The Test Use Project is therefore examining the nature of current assessment practice, the kinds of information such practice yields, the factors influencing the use of the information, the kinds of uses made, and the costs associated with testing. The design of the phase 1 national survey has been influenced by a variety of project planning activities. This report deals with one of these activities--the 1980 reanalysis of test use data that CSE collected in a small-scale study in 1978. The questionnaire distributed to elementary school teachers was based on the premise that teachers' use of tests is influenced by: demographic characteristics of the school; amount and type of mandated testing; and teachers' training and attitudes toward testing and instruction. Results are discussed. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED2055590

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A REANALYSIS OF TEST USE DATA

Jennie P. Yeh

CSE Report No. 143

August, 1980

Test Use Project
Center for the Study of Evaluation
Graduate School of Education, UCLA
Los Angeles, California 90024

TM 810 483

The research reported herein was supported in whole or in part by a grant to the Center for the Study of Evaluation from the National Institute of Education, U. S. Department of Education. However, the opinions and findings expressed here do not necessarily reflect the position or policy of NIE and no official NIE endorsement should be inferred.

Table of Contents

	Page
Abstract	1
Introduction	2
Methods	6
Results	8
Student Demographic Characteristics/Achievement Levels and Test Use	8
Test/Testing Features and Test Use	11
Teacher Training/Knowledge and Test Use	18
Instructional Contingencies and Test Use	29
Discussion	41
Introduction	41
Distance from Information Source and the Need for Information	44
Quality of Information and Risks Associated with Wrong Decisions	46
Testing and Instructional Contingencies	47
Teacher Knowledge and Training	48
References	52

ABSTRACT

This report was developed in CSE's Test Use Project. The general goal of the project is to examine and describe the features and applications of tests and other assessment methods that contribute to the improvement of instruction. The Test Use Project is therefore examining the nature of current assessment practice, the kinds of information such practice yields, the factors influencing the use of the information, the kinds of uses made, and the costs associated with testing.

The first phase of the project (December, 1979 to November, 1981) consists primarily of a national survey of teachers and principals on the kinds of issues suggested above. The second, overlapping phase (February, 1981 to November, 1982) will consist of a small-scale but intensive study of testing costs and the factors identified in phase 1 as influencing the use of test results.

The design of the phase 1 national survey has been influenced by a variety of project planning activities. This report deals with one of these activities --the 1980 reanalysis of test use data that CSE collected in a small-scale study in 1978.

INTRODUCTION

There is little doubt that testing in American schooling is increasing in both scope and visibility. Federal program requirements, school board accountability concerns, national and regional assessment needs, state-mandated minimum competency requirements, and the expansion of curriculum-embedded testing programs have increased the amount of testing. Kirkland (1971) reported that 75 million standardized tests were taken in 1954 by individuals in educational institutions; Goslin (1963) reported that in 1961 the figure had increased to 100 million ability tests per year. Passage of the Elementary and Secondary Education Act of 1966, with its attendant special programs, has increased the amount of standardized testing. Although the exact magnitude of the increase is unknown, we do know that a child takes an average of six full standardized achievement test batteries before he or she graduates from high school (Houts, 1975). We also know (Problems, 1977) that at least 90% of the local education agencies throughout the country administer standardized, norm-referenced tests to children within their purview. In addition, 42 states conduct a state assessment program (Kauffman, 1979) and 37 states have adopted minimum competency legislation (Gorth, 1979); such efforts lead to additional yearly testing for students at various grade levels.

It is equally clear to those who have spent time studying teaching-learning environments, however, that test results seem to be only infrequently used to make decisions about instruction in classrooms. Goslin's (1967) study of testing at the elementary and secondary school level provides

some evidence that teachers used test results primarily to diagnose individual difficulties and to provide information to the student. However, he also reported that the teachers did not rely heavily on this source of information.

The Royal Oak Study (Boyd et al, 1975) also supports the notion that teachers do not rely on required or published test results for instructional decision making. Although teachers reported variable use of results from the district-mandated testing program, there was little evidence that the testing program influenced school curriculum or classroom instruction. For the most part, teachers felt that normed standardized achievement tests were selected by administrators and imposed upon teachers, and did not furnish them any new information. They felt that test results supplied information about students' skills that were already well understood by teachers and parents. Although a small number of teachers thought test results, especially those from criterion-referenced tests, were useful, most felt that the tests given were not useful for planning instruction.

Stetz and Beck (1979) conducted a nationwide study of teachers' opinions of the use and usefulness of standardized tests; their study was conducted in conjunction with the standardization of the Metropolitan Achievement Tests. In the Stetz and Beck study, 80% of the teachers reported making only little use of test data, which is consistent with findings reported by Goslin (1967), and Boyd et al, (1975).

Several explanations support the premise that tests, while common elements in school operations, seem dysfunctional. At the most general level school personnel may be aware that the classroom instruction they

provide has a relatively small influence on children's performance on standardized tests. They may therefore avoid the use of test results that are likely only to underscore and identify educational failures.

Another explanation may reflect the quality of the available instruments (Boyd et al, 1975). In an NIE supported analysis of commercially available measures, researchers at CSE reported that the quality of norm-referenced, standardized achievement tests as well as the newer, criterion-referenced tests, was generally low. (See Hoepfner, 1975; Walker, 1977.) Further, tests might not be used because of the manner in which they report data. Scores are presented in a form convenient for decisions about groups rather than about individual learners; teachers' decisions, which focus in large part on the successes and failures of individual children, require a different kind of information. (See Lortie, 1975.)

Another factor contributing to limited use of tests may stem from teachers' understanding of the design, interpretation, and uses of tests for decision making (Cramer & Slakter, 1968). Common perceptions about commercial test agencies may be that while the agencies "know what they're doing," the technical processes involved in test development are so arcane that it is difficult for a teacher to know if he or she is using and interpreting the test properly. Such perceptions may be underscored by the level of training that teachers, in their certification sequences, receive in test development and interpretation. Training is often limited to interpretation of IQ or stanine scores, without much treatment of the advantages and limitations of different kinds of tests. These inadequacies in teacher training are not surprising; those who provide such training rarely have strong backgrounds in the field of measurement.

A final potential reason for the lack of test use may reside in contextual variables associated with individual schools. For instance, the number of instructional alternatives available in a school setting may influence the amount of testing required. Because there are a limited number of instructional alternatives in their classrooms, and thus only a limited range of decisions that can be made, teachers may feel that those decisions can be made on the basis of information more readily available to them than test results. Another contextual variable might reflect teachers' belief that test results provide minimal help in understanding the children in their classrooms.

METHODS

Since previous studies (Goslin, 1965, 1967) indicated that teachers' testing practices varied with school location and parents' SES, we decided, in 1978, to select a sample of schools that covered a wide range of locations and SES groups. Five districts in California were selected, one suburban, one rural, and three urban, so as to represent a wide range of economic and ethnic composition. Within each district, four elementary schools were randomly selected for a total of 20 schools. The factors considered in school selection were (1) enrollment size, since we wanted to insure that both large and small schools were represented, and (2) the availability of compensatory educational funding. Such funding is an important factor, first, because while the major compensatory educational programs available to California schools require student achievement test results for evaluation or accountability purposes, a test required for one program can seldom be used to fulfill the requirements of another program. Therefore, students in schools receiving multiple fundings take a greater number of mandated tests. A second reason for using compensatory educational funding as a sampling factor was suggested by Baker's (1976) finding that teachers tend to differ in their reliance on test results depending upon whether their school receives compensatory funding. Teachers in schools with compensatory education funding reported less use of standardized test results and more use of curriculum-embedded and teacher-constructed tests for diagnostic purposes in instruction.

We developed our teacher questionnaire for the 1978 study on the premise that teachers' use of tests is influenced by the following factors:

1. Demographic charcteristics of the school/classroom
2. Amount and type of mandated testing; characteristics of the tests
3. Teachers' training, experience, and attitudes toward testing; toward instruction.

The ensuing ten-page questionnaire consisted of the following sections:

Section A: Classroom organization and instructional practices

B: Methods teachers use to assess programs

C: Use and understanding of mandated tests

D: Opinions on currently used assessment methods

E: Teachers' experience and training

Additional information on the schools' socioeconomic index (SES), percent of students receiving aid to families with dependent children (AFDC), and the percentile rank of the state assessment score was obtained directly from existing data files.

Questionnaires were mailed to principals, who were asked to distribute them to all kindergarten through sixth-grade teachers at each school. All twenty schools from the five districts received the questionnaire. However, one school was not able to participate because it was taking part in an experimental busing program that involved both students and teachers.

Two-hundred and sixty questionnaires were received; we estimate that the return rate was approximately sixty percent.

RESULTS

STUDENT DEMOGRAPHIC CHARACTERISTICS/ACHIEVEMENT LEVELS AND TEST USE

To describe the school features--student demographic characteristics and achievement levels--presumed to affect test performance and test use, the study employed two variables. These were (1) a measure of socioeconomic status, which was the percent of families receiving aid for dependent children, and (2) the average third-grade percentile score on the California State Assessment test for the school. A high, significant correlation (.82) between the two variables for our sample indicates that the lower the income level of the school population (represented by a higher percent of AFDC), the lower the achievement level for that school.

As early as 1923, researchers attempted to find explanations for the relationships between SES and achievement. Hypotheses range from the extreme, e.g., that minorities' achievement can be explained in genetic terms, to the more plausible, e.g., that the measures used to assess intelligence or achievement are culturally biased. Cultural bias has been discussed extensively by a number of authors, particularly in terms of the linguistic bias in standardized tests. Gardner and Gardner (1978) and Hilliard (1979) feel that test content is invalid for Black students with the result that reading and linguistic skills are erroneously tested which, they surmise, results in spuriously low scores in achievement. Christiansen and Livermore (1970) examined the influence of social class on IQ scores. They compared the performance of Anglo-American and Spanish-American students from low and middle social classes on the WISC, and concluded that the Hispanic

children scored lower because of an English fluency deficiency rather than because of innate intellectual capacity. Green (1973) feels that cultural differences in cognitive style, values, and language use between test developers and examinees may result in the examinees misunderstanding what the test presents and what is expected of them.

In an examination of other variables that affect achievement levels, Mercer (1971, 1972, 1973) compared intelligence test scores, measures of adaptive behavior, and sociocultural variables among Anglo, Black, and Chicano subjects whose IQ scores were below average (less than 85). She found that with regard to intelligence test scores, the tests are valid for Anglos, but not for Blacks or Chicanos. In her article concerning the effects of labelling the mentally retarded (1972) she states, "... What the IQ test measures, to a significant extent, is the child's exposure to Anglo culture. The more 'Anglicized' a non-Anglo child is, the better he does on the IQ test" (Mercer, 1972, p. 95).

Mayeske (1971) drew similar conclusions after reanalyzing sixth-grade students' achievement scores. He found that the differences in academic achievement performance among six racial/ethnic groups (Indian, Mexican American, Puerto Rican, Black, Asian, and White) are reduced when certain sociocultural variables are corrected for statistically. These variables included: social and economic well-being of the family; the presence or absence of key family members; the student's and parents' aspirations for his/her schooling; their beliefs about how the person might benefit from an education; the activities they engaged in to support these aspirations; one's region of residence; and the achievement and motivational levels the student takes to school with him/her.

Before the scores were adjusted (with a distribution mean of 50), Whites attained the highest score with Asians following them by about 4 points. Approximately 5 to 7 points below the Asians were the Indians, Mexican Americans, and Blacks. The Puerto Ricans trailed these groups by another 4 points. Mayeske concluded from his reanalysis that "no inferences can be made about the 'independent effect' of membership in a particular racial ethnic group on academic achievement" (p. 21) because the score a student obtains is almost completely confounded with many social conditions.

To the extent that minority student populations are associated with lower socioeconomic status schools, and since achievement is determined by standardized tests that may be biased against the background and experience of minority populations, it is not surprising that lower SES schools are also low achieving schools. It appears, further, as one would expect, that teachers in these schools are also aware of the limitations of most standardized tests in describing the ability or performance levels of their students.

Data from teachers in our 1978 survey support this expectation. Higher achieving schools, i.e., those whose students were able to score higher relative to the mainstream culture norms, tended to report more positive attitudes about required tests and about the time spent on them. The teachers thought that IQ and students' true reading or math ability significantly influenced scores on required tests. On the other hand, teachers at lower achieving schools tended to have more positive attitudes about developing their own tests and negative attitudes about required testing; they would react positively if required testing was eliminated at their schools.

The lower achieving schools also felt that unusual circumstances had a great influence on the scores on required tests. Low-SES/low-achieving schools, further, might use test results for initial placement of their students in reading and math. However, in their math programs, they do not use tests for continual student assessment throughout the school year; they may use tests in this manner in their reading programs. The low-achieving schools also tended not to use the test results for communicating students' progress to parents, in contrast to the higher SES schools who tended to use test results for this purpose.

TEST/TESTING FEATURES AND TEST USE

Teachers' commonly used methods of assessment are limited in number and precision; essentially, teachers rely upon results from achievement tests or on observations of their students (Lortie, 1975). The tests teachers give cover a range of formality including the very formal standardized norm-referenced tests, designed to differentiate among examinees: criterion-referenced tests designed to assess student mastery of a subject domain; curriculum-specific tests accompanying and/or referenced to commercially available text series; teacher-made tests; and the less formal, such as work samples and short oral quizzes.

These available testing techniques vary in amount of teacher and student time required, ease of administration, scoring and interpretation, and reliability and quality of information provided (Klein, 1970; Howe, 1978a, 1978b; Kahn, 1978; Broekhoff, 1978; Ahmann, 197[illegible]). Given this available variety of tests and their strengths and weakness, our study

attempted to determine teachers' preference and use for classroom decision-making.

Types of Tests Teachers Give to Students

According to our 1978 survey, teachers differ (as will be discussed later) in the degree to which they give periodic tests in math and reading. But of all the tests they gave to their students, curriculum-embedded and teacher-made tests constituted the greatest proportion. Locally-developed tests (district or school level) constituted a moderate proportion and standardized tests constituted the smallest proportion of all the types of tests teachers administered.

Range of Test Information Used by Teachers

For making decisions for initial instructional placement of students, a greater percentage (58% for reading and 65.8% for mathematics) of teachers reported relying upon test results rather than other sources of information such as informal assessment or students' progress in a textbook. Test results seem to provide the teacher with a quick and acceptable initial estimate of ability for those students with whom they are unfamiliar, e.g., at the beginning of the school year.

However, for assessing student progress throughout the school year, teachers reported that of the several alternative sources of information, they most frequently relied upon interactions with or observations of students, informal oral quizzes, or teacher-made tests. Standardized tests were the least frequently used sources of information for ongoing

assessment of student progress. Apparently, once initial placements are made and teachers have access to a number of information sources, their reliance on formal test results decreases.

Obviously, teachers' in-class interactions, observations, and teacher-made assessment techniques offer a greater wealth of information about students. Further, the content, format, and timing of this information is more immediately relevant to instruction than the information from measures developed outside the classroom and designed to "cover" a range of possible grade level instructional content and method. It seems only reasonable, then, that teachers rely upon the apparently more relevant measures to a greater degree when they are available, i.e., after the school year is underway. The implicit instructional value of these measures may explain why, as a group, they were cited as the most frequently used assessment methods. Within this group of measures, teachers most frequently used information from their own tests for the following purposes: making instructional decisions about students; evaluating the effectiveness of teaching methods and materials; and providing information to parents and/or staff. The least cited use of information from teacher-made tests was for assigning grades, although this purpose ranked between "sometimes used" and "frequently used."

The most frequently reported use of test results from mandated assessments was communication with parents and/or staff and for evaluating the effectiveness of teaching methods and materials. Required test results, usually from norm-referenced tests, seemed to function as a standard for comparison. These less student-direct uses of required tests contrast with

how teachers use their own test results, i.e., for instructional decisions with greater impact on students.

Quality Concerns of Currently Used Tests

Despite the busy schedules of the elementary school teachers, we found that they constructed 50% of the tests used in the classroom. Also, despite the large number of ready-made tests available, 53% of the teachers reported that they regularly construct tests for classroom use. The most frequently cited reasons for this low reliance on available resources reflect criticisms of commercially available tests: lack of reliability and validity of both norm-referenced and criterion-referenced tests (Petrosko, 1978; Crew & Whitney, 1978); cultural bias (Gardner, 1977, 1978); lack of match between test items and instructional content; lack of diagnostic information; cost for test purchase and student time; timeliness of testing and reporting.

Teachers in our survey reported that the most important reasons for developing their own tests are (1) these tests more accurately assess the effects of their instruction, and (2) the wording and format are most suitable to their own students (Table 1).

TABLE 1

Teachers' Reasons for Developing Their Own Tests

(1 = important 5 = very important)

(N = 210 Teachers)

Reasons	Average Rating	Standard Deviation
a. Funds are not available for buying tests.	1.8	1.30
b. My own tests have more suitable wording and format for my students.	4.5	1.06
c. There is not time to order published tests.	1.6	1.16
d. Information about prepared tests is not readily available.	1.9	1.20
e. My own tests more accurately assess the effects of my instruction.	4.6	.89

When choosing from available tests, the qualities teachers considered most important are (1) clear format, (2) similarity to class materials, and (3) accurate prediction of student achievement (Table 2). Clarity of format was considered to be more important by lower (k-4) grade teachers ($\bar{x}$ = 4.7) than upper grade teachers ($\bar{x}$ = 4.3).

TABLE 2

Qualities Teachers Consider When Selecting Tests

(1 = important 5 = very important)

(N = 210 Teachers)

Factors of Concern	Average Rating	Standard Deviation
a. The test material is similar to what I present in class.	4.5	.88
b. The test is simple to administer and/or score.	4.2	1.05
c. The test has clear format, pictures, directions.	4.6	.70
d. The test publisher has a good reputation.	2.8	1.46
e. The test accurately predicts student achievement.	4.4	.97

Commercially produced curriculum materials presently structure much of the classroom instructional content. With the advent of a behavioristic approach to learning and instruction (Skinner, 1968), the role of testing, referenced to curriculum materials, began to assume great importance. Our study indicates that these curriculum-embedded tests constitute almost half of the tests teachers give to students in their classrooms. How do teachers actually use these tests? How well are these tests constructed? How do teachers feel about the quality, administration, and interpretation of this type of test? These are important, though as yet unexplored, questions about the linkages between testing and instruction.

Curriculum-embedded tests were originally designed to serve both as

diagnostic instruments for determining students' progress through program outcomes and as devices for formative evaluation and revision of the program materials. In fact, the very specific match required between instructional objectives and test items for developing validated instructional programs led to the birth of criterion-referenced testing (Glass, 1980). However, due to the newness of curriculum-embedded tests, generally little is known about this quality.

Our finding that teachers do distinguish features affecting quality of required and teacher-made tests corresponds with the distinctions teachers make in the use of test data for instructional decisions about students (such as grouping, reteaching), evaluating teaching methods and materials, assigning grades, and providing information to parents and staff. In general, however, teachers reported greater use of results from teacher-made tests and less reliance on results from required tests (Table 3).

TABLE 3

How Teachers Use Information From Their Own Tests vs. Required Tests

(1 = never 4 = always)

(N = 210 Teachers)

Use of Information	Teachers' Own Tests		Required Tests	
	$\bar{x}$	σ	$\bar{x}$	σ
a. Making instructional decisions (e.g., grouping; progressing to the next unit)	3.2	.77	1.8	.80
b. Evaluating the effectiveness of the teaching methods or materials	3.1	.85	2.0	.85
c. Assigning grades	2.6	1.01	1.4	.63
d. Providing information to parents and/or staff (other teachers or administrators)	3.0	.83	2.1	.82

TEACHER TRAINING/KNOWLEDGE AND TEST USE

The literature review presents a variety of issues relating to teacher knowledge of tests and testing, and teacher use of test results. Much of the literature either explicitly cites lack of teacher knowledge of tests and testing or, by virtue of the subject matter presented, implicitly highlights such lack of knowledge.

Goslin (1967) found: (1) less than 40% of all teachers have had minimal formal training (one course) in test and measurement techniques; yet large numbers of teachers, especially in the elementary grades, are responsible for administering standardized achievement tests; (2) teachers

tend to view standardized tests as relatively accurate measures of student achievement, and to see the abilities measured by these tests as important determinants of academic success; but (3) teachers indicated a rather low degree of test use in terms of grading and advising pupils and in providing them with feedback (the higher the degree of teacher training on testing, the higher the use of test scores).

Hastings, Runkel, and Damrin (1961) also agree that test use depends on teacher knowledge of tests and their interpretation. This belief is implicit in a number of works (e.g., Gorow, 1966) attempting to provide teachers with information on how to design teacher-made test items and how to improve tests through analysis of test results, and evidenced by the numerous "handbooks" like Bauerfeind's (1963) work on basic issues in testing such as validity and reliability and the building blocks of an effective school testing program.

Data from our 1978 study indicate teachers' background knowledge in testing is, indeed, limited (Table 4). Although a majority of our teacher sample reported units beyond their bachelor's degree and credential, 61.4% indicated only one college course in testing. And this situation is not remedied via district in-service programs; 76.8% of the teachers reported only one or no in-service course on testing.

In addition to determining the amount of training teachers have or should have in testing theory, more immediately interesting questions arise: Are teachers capable of interpreting the test results commonly provided to them? In what area(s) do they want additional training?

TABLE 4

Teacher Training and Knowledge of Testing

Frequencies and Percent

Educational level	N	Percent
BA	8	3.1
BA and teaching credential	54	21.3
BA and units toward MA	100	39.4
MA	57	22.4
MA and units toward Ph.D.	33	13.0
Ph.D. or Ed.D	2	.8

Knowledge of Grade Equivalence

	N	%
incorrect	120	50.4
correct	118	49.6

College Courses in Testing

Number of courses	N	Percent
0 courses	54	23.2
1 course	89	38.2
2 courses	52	22.3
3 courses	38	16.3

In-service Courses in Testing

0 courses	54	43.2
1 course	42	33.6
2 courses +	29	23.2

Knowledge of Percentile Score

	N	%
incorrect	87	36.1
correct	154	63.9

Our original plan to give teachers a quiz on test knowledge along with the questionnaire was field-tested but proven to be infeasible. Since standardized test results are commonly reported in the form of percentiles and grade equivalent scores, two multiple-choice questions were used to ask teachers about their understanding of the meaning of "percentile" and "grade equivalence."

1. If one of your students scored at the 80th percentile in math and the 70th percentile in reading on published tests, how would you interpret these results?
2. If one of your students received a grade-equivalent score two years higher than his/her actual grade level, how would you interpret these results?

58.5% of the teachers answered item 1 and 45% answered item 2 correctly. The fact of having had more classes in testing did not improve a teacher's chances of answering these items correctly ($r = .14$). However, when asked to rank-order five in-service topics according to their "helpfulness," interpreting results from standardized tests ranked third after using test results in instructional planning and selecting tests for the classroom. Constructing tests ranked fourth and using criterion-referenced tests, a least helpful fifth.

Although how to construct tests is not high on the teachers' lists of things they need to learn, the quality of teacher-made tests has been questioned. Leiter (1976) suggested that the background most teachers have in tests and testing will lead to the development of unreliable tests. Ebel (1967) described the problems in teacher-constructed tests as: too

heavy reliance on subjective evaluation; leaving testing too late to be of instructional use, and developing tests that do not sufficiently sample student knowledge and ability in a given curricular area. More specifically, he cites development of trivial and ambiguous items; lack of teacher knowledge of the measurement errors to which tests are subject, and failure to test the effectiveness of their tests by statistical analysis of results.

Besides the teachers' training and knowledge in testing, another factor that may influence teachers' test use is their feeling toward testing (Boyd, McKenna, Stake & Yachinsky, 1975).

When considering testing cost in terms of time spent in required testing vs. the benefits that accrued, teachers in our sample felt that the amount was between "about right" and "slightly too much" (average 2.5 on a 1-5 scale with 1 being "too much"). Responses differed significantly depending on teachers' experience level. Teachers with less than 16 years experience were more likely to respond that "too much" time (2.2) was spent in testing in relation to the benefits received, than were teachers with at least 16 years of experience (2.6). This finding may reflect less experienced teachers' dissatisfaction with the actual tests or more experienced teachers' acceptance of their testing program.

Teachers indicated that if required testing were to be discontinued, their reaction would be favorable (3.8 on a 1-5 scale, with 1 being "with disapproval" and 5 being "with approval"), especially primary grade teachers whose average rating was 4.0. Teachers generally perceived that administrators would be less positive ($\bar{x}$= 2.9) and parents the least positive ($\bar{x}$= 2.5) about eliminating required testing.

The responses to these two questions, and the fact that teachers reported low use of required test results for instructional purposes, suggest that although teachers did not voice strong opposition to current testing requirements, their attitude is certainly very reserved.

What explains teachers' attitudes toward and use of required testing results? Teachers surveyed in the Royal Oaks study (Boyd et al., 1975) felt classroom instruction has slightly higher than median level of effect on test scores (3.7 on a 5 point scale with 5 being "great effect"). But they rated the quality of required tests (directions, content, format, physical characteristics), student test-taking skills, student motivation, and unusual circumstances as having more influence on students' test scores than students' actual reading and math ability, IQ, or the quality of instruction (Table 5).

TABLE 5

Teachers' Opinions of Factors Which May Influence Student Test Scores on Required Tests

(1=little influence, 5=great influence)

(N = 220)

a.	Students' test-taking skills	4.4	.74
b.	Quality of classroom instruction	3.7	.96
c.	Student IQ	3.7	.97
c.	Student motivation	4.3	.77
e.	Parent interest	3.0	1.77
f.	Quality of test by directions, content, answer format, physical characteristics)	4.4	.79
g.	Unusual circumstances (special activities, distractions)	4.2	.90
h.	Students' true reading and math ability	3.9	.91

Teachers' lack of training in testing and their limited knowledge of how to interpret test results are frequently cited as principal reasons for teachers' low (or non-existent) use of information from required tests. We therefore decided to examine the relationships among teacher knowledge of tests and teacher use of tests on the basis of (1) self-reported levels of training in tests and testing, (2) years of teaching experience, and (3) actual ability to interpret test scores. We examined these relationships by using a latent structure model, where ξ_1 is the latent variable of student/school characteristics as estimated by SES and the school's achievement test score from CAP; ξ_2 is the amount of teacher training estimated by both college courses and in-service sessions in tests and testing; ξ_3 is teacher experience measured by the number of years teaching. N_1 represents teachers' ability to interpret test scores, and N_2 is teachers' reported use of test results.

As discussed in the section on test characteristics, teachers tend to differentially use results from required testing for communication/support and for instructional decisions. These two users were tested by using the latent structure model. In this model, relationships between the variables of interest are translated into causal paths and tested for significance via the LISREL computer program (Jöreskog & Sörbom, 1978). The program provides estimates of structural relations among latent variables based on observed covariances between variables selected as indicators of the latent constructs. Figure 1 illustrates the structural model of teacher training/knowledge and the use of test results for communication/support.

Figure 2 is the structural model for teacher training/knowledge and use of test results for instructional decision making. The estimated structural equation parameters are given above each causal path with their t-values in parentheses.

Figure 1 indicates that in low SES schools where student achievement is also low, teachers are less likely to use test results for communication/support purpose (γ= -.41). As discussed earlier, there are a number of possible reasons for this finding: teachers in low SES/low achieving schools may feel test content does not reflect classroom instructional content, or that factors unrelated to student knowledge have greater influence over test results. Also, teachers in these schools may not use test results in communicating with parents because parents show less interest (Yeh, 1978, teacher interview data).

The direct causal path between the amount of training the teacher had in tests/testing and his/her test use indicates that the more training a teacher has had, the more likely he/she is to use test results (r=.25). However, given the above relationships, the path from teacher experience, teacher knowledge, and test use suggests that the more experienced teachers are better able to interpret frequently used methods of test reporting (percentile, grade equivalence), but that this knowledge seems to inhibit teachers from using results from required tests for communication/support purposes. A causal path from teacher training (ξ_2) to teacher knowledge (n_1) was included in the original model because it seemed reasonable to assume that teachers with more training had more knowledge in test interpretation. Based on the model, the path coefficient was estimated to be

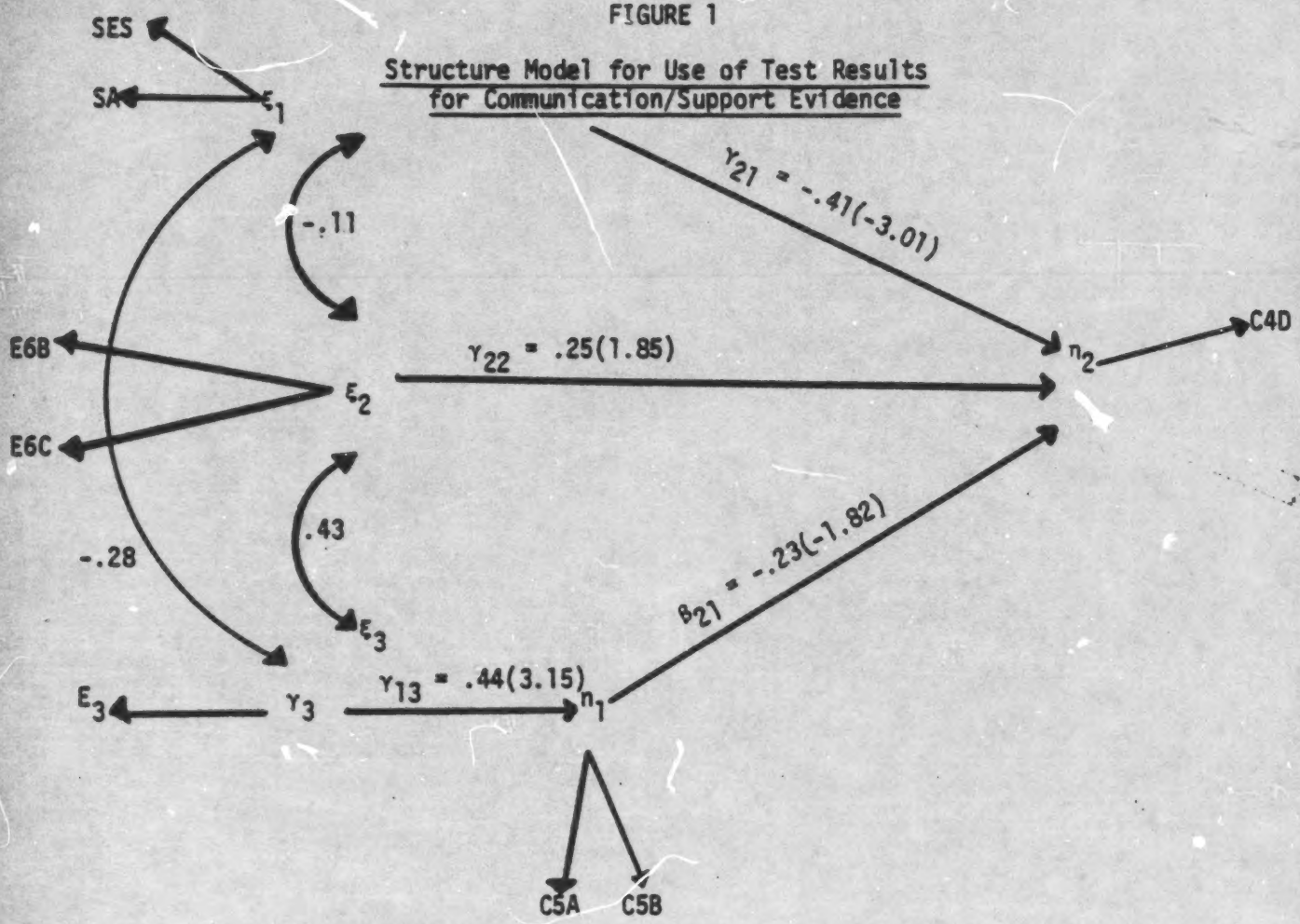
-.08, which was found to be non-significant. A number of reasons are possible for the failure to obtain this relationship: one is that teacher training does not teach in detail how various norms are established; another is that our items used to assess teacher knowledge did not adequately reflect teachers' knowledge in test interpretation.

Another path included in the original model was between teaching experience and test use (ξ_2). Although teachers as a group reported that they did not use required test results frequently (2.0 on a scale of 4, where 1 is "never" and 4 is "always"), more experienced teachers tend to report higher use (1.5 for teachers with 1-7 years of experience, 1.8 for teachers with 8-15 years of experience, and 1.9 for teachers with more than 16 years teaching experience). Based on the model, the path coefficient was estimated to be .02 and statistically non-significant. Therefore, it suggests the relationship between teaching experience and test use is an artifact of experienced teachers. That is, experienced teachers tend to have had more training in testing and these more fully trained teachers tend to use test results more for communication/support.

The model in Figure 2 examines the relationship between teacher knowledge (N_1) and use of required test results for instructional purposes (N_3). The overall model fit was quite good, as indicated by a chi-square goodness of fit of x^2=12.9 with df=17, p=.74. This represents a significant improvement over the measurement model: x^2=41.32 with df=25 and p=.02, wherein the structural relations are set to zero; and over the null model: x^2=249.9 with df=28, p=.00, where the number of observed variables is equal to the number of latent constructs, and all measurement error variances are

FIGURE 1

Structure Model for Use of Test Results for Communication/Support Evidence

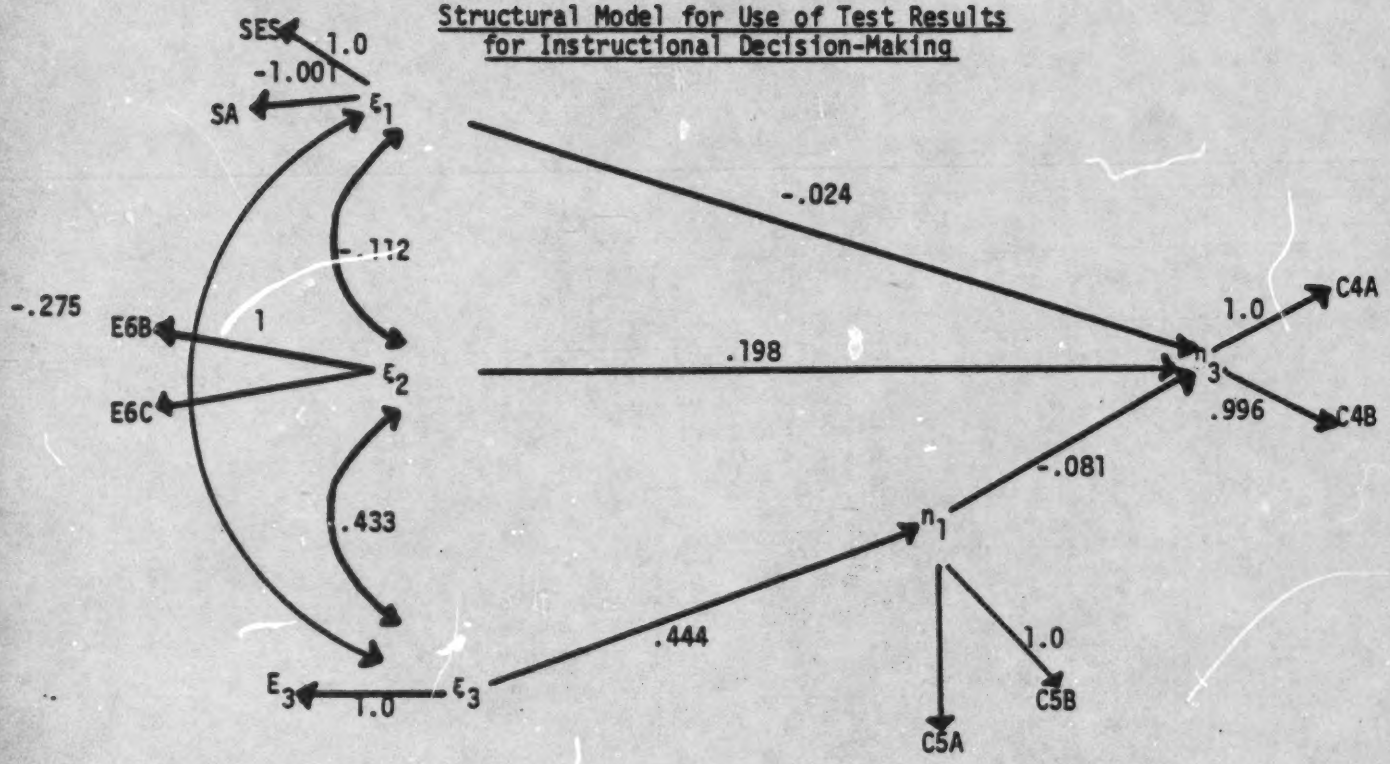


Notation Key

SES = socio-economic status
SA = school achievement on state assessment (CAP)
E6B = college courses in testing
E6C = in-service courses in testing
E3 = teaching experience (no. of years)
C5A = knowledge of percentile interpretation
C5B = knowledge of grade equivalency interpretation
C4D = providing information to others (parents/staff)

FIGURE 2

Structural Model for Use of Test Results for Instructional Decision-Making



Notation Key

SES = socio-economic status

SA = school achievement on state assessment (CAP)

E6B = college courses in testing

E6C = in-service courses in testing

E3 = teaching experience (no. of years)

C5A = knowledge of percentile interpretation

C5B = knowledge of grade equivalency interpretation

C4A = making instructional decisions about students

C4B = making instructional decisions about (evaluating) teaching materials/methods

TABLE 6

Direct and Indirect Effects on Use of Test Results for Communication/Support Evidence by Hypothesized Causes

Causes	Direct Effects	Total Indirect Effects	Total Effects of Hypothesized Causes
Test Knowledge			
Teaching Experience	.440		.440
Use of test results to inform parents			
Test Knowledge	.230	.110	-.120
School SES Level	-.41	-.024	-.434
Teacher's Testing Courses	.25	.048	.298

TABLE 7

Direct and Indirect Effects on Use of Test Results for Instructional Decision Making by Hypothesized Causes

Causes	Direct Effects	Total Indirect Effects	Total Effects of Hypothesized Causes
Test Knowledge			
Teaching Experience	.44		.44
Use of test results to make instructional decisions			
Test Knowledge	.081	.044	-.037
School SES Level	.024	-.007	-.031
Teacher's Testing Courses	.198	-.011	.187

set to zero. However, none of the causal paths leading to N_3 was statistically significant. Therefore, our data suggest that teachers' training in testing, their ability to interpret test scores, and their teaching experience are not factors affecting use of test results for making instruction-related decisions.

INSTRUCTIONAL CONTINGENCIES AND TEST USE

According to the decision-making model (Shavelson & Borko, 1979; Airasian, Kellaghan, Madaus, & Pedulla, 1977), teaching is described as a rational, information-based process in which testing is viewed as an information-gathering activity. When a teacher is planning for instruction, she/he is assumed to have a number of available teaching materials and/or methods which can help students attain some goal. To choose among the alternatives, a teacher needs to gather information about the students in order to match instruction to the students' needs.

The information-gathering techniques available to teachers are limited to formal and informal tests, or observations of and interactions with students. This distinction has been described in terms of products and learning processes. The range of product assessments include the very formal, standardized norm- or criterion-referenced tests, curriculum-embedded tests, teacher made tests, and less formal samples and short oral quizzes. Process assessment includes observation and interaction. These available techniques vary in the amount of teacher and student time

required, ease of administration and interpretation, reliability, validity, and quality of information provided. Given the amount and type of information sources available, a teacher chooses to use or not use information to guide what she/he judges to be sound instructional decisions. Factors that are likely to influence teachers' information sources and uses are those affecting the amount of time available for the teacher to gather direct assessment data through interactions with or observations of each individual child. For example, it is probably more likely for teachers to use information gathered from a variety of sources when there is a greater range of available instructional options such as aides, remedial/enrichment resources, and materials. The following sections report our 1978 data on such factors influencing test use for instructional decision-making.

Source of Information Used for Initial Placement

Teachers reported that compared to other sources of information, they most frequently used some kind of formal test results (from commercially published or teacher-made tests) to assess student skills in reading and math at the beginning of the school year. Fifty-eight percent reported using test results most often for initial reading placement and 65.8% reported using test results most often for initial math placement. The second most common method for student placement, informal assessment techniques (e.g., teacher observation, student work samples), was more frequently used for reading placement (24.2%) than for math placement (16.9%). Test results seem to provide the teacher with a quick and acceptable estimate of ability of those students with whom the teacher is unfamiliar (Table 8).

TABLE 8

Sources of Information Teachers Use to Place Students In Reading and Math at the Beginning of the Year

	Reading Placement	Math Placement
A. Test results (e.g., standardized, teacher made)	58.7%(148)	65.8%(160)
B. Previous placement	15.1 (40)	16.0 (39)
C. Informal assessment (e.g., teacher observation, work sample)	24.2 (61)	16.9 (41)
D. Other	1.2 (3)	1.2 (3)

Sources of Information Used for Assessment Throughout the Year

Teachers reported that after they had made initial placement decisions, they most frequently assessed their students throughout the year on the basis of information from interactions with or observations of students, the results of informal assessment techniques (e.g., oral quizzes, reading aloud) or the results from teacher-developed tests. The least frequently used sources of information were the results from standardized and instructional program or curriculum-embedded tests, while moderate use was made of information about students' progress, placement in a book, and work assignments. Once initial placements are made, teachers make less use of standardized or instructional program tests.

Teachers' use of these assessment methods differed with the presence of an aide. Teachers with aides were more likely to use information from a variety of sources. They were more likely to use the results from instructional program tests, students' progress or placement in a book, and the results from informal assessment techniques as well as the results

of interaction with or observation of students. Although aides and curriculum-embedded tests may both be required by a funded program, it seems plausible that with additional help in the classroom, the teacher has more time for personal involvement with students and also more time to use available information (Table 9).

TABLE 9

Sources of Information Teachers Use to Assess Student Progress Throughout the Year

(1=never use, 5=always use)

	Any paid aides	No paid aides	Some paid aides	TOTAL SAMPLE
A Standardized test results		2.3(69)	2.2(151)	2.2(221)
B Instructional program test results		*2.8(67)	3.2(146)	3.1(215)
C Teacher-developed test results		3.9(71)	3.9(153)	4.0(226)
D Student's place in book		Δ3.5(72)	3.8(150)	3.7(224)
E Work assignments		3.8(72)	3.7(148)	3.7(222)
F Interaction with student		4.2(72)	4.3(154)	4.2(227)
G Informal assessment		*3.8(71)	4.1(151)	4.0(225)

* = one-way ANOVA significant at $\alpha < = .05$

Δ = one-way ANOVA significant at $.05 < \alpha \leq = .10$

Frequency of Math and Reading Tests

In general, teachers at every grade level reported giving math tests more often than reading tests. 29.5% of the teachers reported giving monthly reading tests, 29.5% reported giving these tests weekly, while 8.7% reported giving daily reading tests. This compares to 18.1% of the teachers who reported giving monthly math tests, 50% who reported giving weekly math tests, and 9.7% who reported giving daily math tests. 32.3% of the teachers gave less than 1-2 reading tests a semester, whereas 21.8% gave less than 1-2 math tests a semester. This finding may reflect the fact that more drill and practice is necessary in math and/or that math tests are easier to construct, score, and interpret than reading tests (Table 10).

There was also a difference in teachers' testing frequency depending on the grade level taught. The results showed a trend for upper-grade elementary teachers to give tests more often than primary grade teachers. (These results seem to confirm information from the interviews, Yeh, 1978, that teachers gave fewer tests to younger children because they felt these students lacked the powers of concentration and skills necessary for taking tests.) But when primary grade teachers did give instructional program tests, they used the results to assess student progress throughout the year.

TABLE 10

How Often Teachers Give Reading and Math Tests
(% of Teachers)

	never	1-2 times a year	1-2 times a semester	monthly	weekly	daily	Median	N
a. Reading tests	4.5%	11.2%	17.0%	29.5%	29.5%	8.7%	3.1	213
b. Math tests	3.8%	8.8%	9.2%	18.1%	50.0%	9.7%	3.7	210

Types of Tests Administered

Instructional program and teacher-developed tests constituted the greatest proportion of all tests teachers administered to their students. The finding concerning teacher-developed tests corresponds to the previous finding that results from these tests were most frequently used to assess student progress throughout the year. Whereas a great proportion of instructional program tests were administered, these tests were least frequently used throughout the year. Locally-developed tests (district or school level) constituted a moderate proportion and standardized tests the smallest proportion of all the types of tests teachers administered. Teachers with paid aides and teachers in team-taught classrooms tended to give a large proportion of tests that were locally developed by someone other than themselves (e.g., district tests) than did teachers without aides and teachers in self-contained classrooms. It may be that these tests have been developed by the aides of the team teachers and simply administered by the respondent teacher. Team teachers gave a greater proportion of all types of tests than did teachers in self-contained classrooms, although there was no significant difference in their frequency of testing (Table 11). As hypothesized earlier, the classrooms with aides are likely to be part of funded programs which may require district- or school-developed tests. This would account for the administration of a greater proportion of these types of tests.

TABLE 11

Teachers' Use of Different Types of Tests

(1=none, 6=all)

	A Standardized Tests	B Instructional Program Tests	C Locally Developed Tests	D Teacher-Developed Tests
Any Paid Aides				
no paid aides	2.3(63)	3.6(68)	*2.2(62)	3.5(69)
some paid aides	2.1(134)	3.6(143)	2.8(126)	3.5(134)
Classroom Organization				
team-taught classroom	2.3(71)	3.8(77)	*2.9(71)	3.6(125)
self-contained classroom	2.1(119)	3.6(127)	2.4(111)	3.6(69)
TOTAL SAMPLE	2.2(197)	3.6(211)	2.6(188)	3.5(203)

*one-way analysis of variance significant at $\alpha \leq .05$

A review of the observed relationships between instructional contingencies and test use discussed in the previous section suggests that:

1) The more closely a test is related to instruction, the more likely it will be used.
2) The more individualized instruction becomes, the less likely that results from a standarized test will be used.
3) Given the busy work schedules of most teachers, the ready availability of a test is a prime concern in frequency of testing.
4) Individualization of instruction seems to be a function of the availability of instructional resources such as aides, reading specialists, etc.

These results, in turn, are correlated with other variables such as SES, test validity, teachers' instructional practices, and teachers' attitudes toward required testing in particular and testing in general. For example, within a district, lower SES schools receive more compensatory educational funding which can be used for additional teaching personnel such as aides or specialists which in turn enable the teacher to individualize instruction. On the other hand, in the schools studied, compensatory educational funding, as a rule, required testing for accountability. These schools are therefore required to give a number of additional mandated tests. This may influence teachers to use test results more because the information is available or more salient to them.

However, the amount of required testing and test use is <u>not</u> linearly

correlated, perhaps because the information provided by the various tests is redundant. Another data-supported explanation is that teachers ignore test results because the standardized, norm-referenced tests currently required are not relevant or sensitive to classroom instructional content. Based on the literature previously cited on test bias, we suspect this phenomenon to be especially true in the low SES schools.

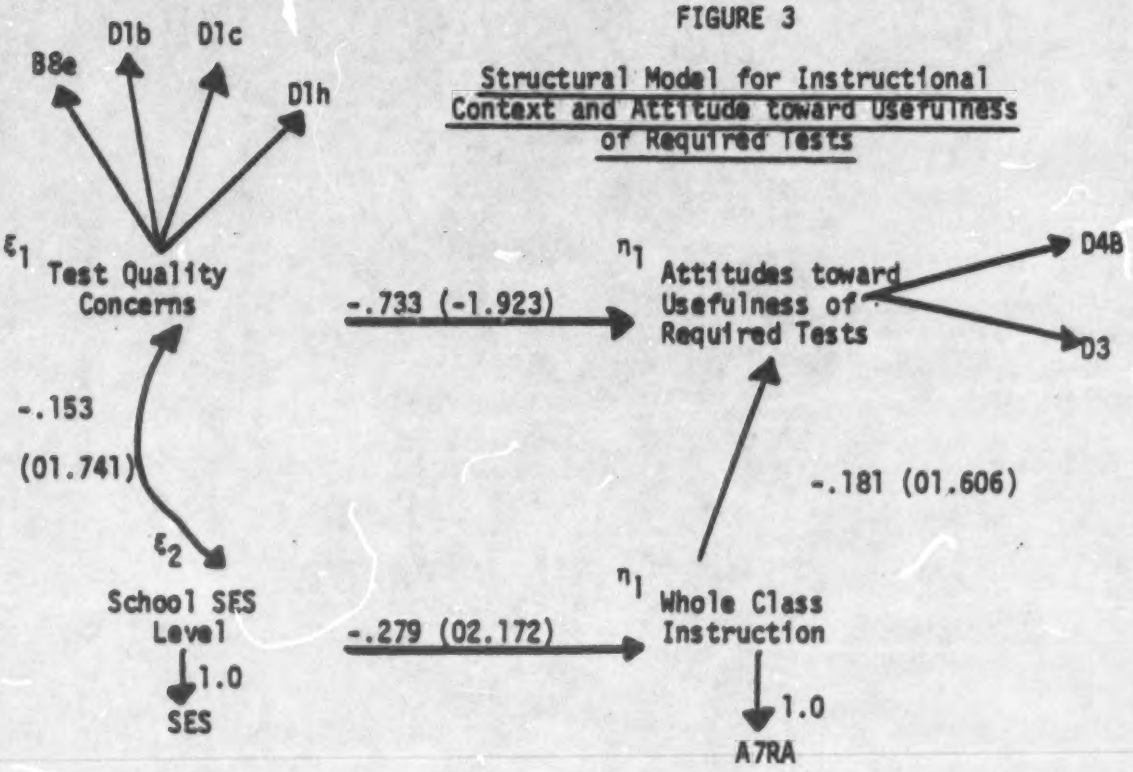
The structural model was again used to account for some of the observed correlations between variables on the questionnaire concerning attitudes toward use of information available from required tests. In this model (Figure 3), the relations described between school SES, general test quality concerns, and attitudes toward the usefulness of required tests are translated into causal paths and tested for significance via the LISREL computer program (Jöreskog & Sörbom, 1978) previously described.

The construct of Test Quality Concerns includes observed variables: the teacher's desire that the test accurately predict student achievement and the teacher's belief that quality of classroom instruction, student IQ, and student's true reading and math ability influence student performance on required tests. All these variables were perceived to indicate concerns about the ability of a test to detect relevant and reliable ability differences due to a particular testing situation.

The construct of school SES level was indicated by the percentile of families receiving aid for dependent children--the higher the aid level, the lower the SES. For whole class instruction, the variable of percent of time spent teaching reading to the whole class was used. Finally,

FIGURE 3

Structural Model for Instructional Context and Attitude toward Usefulness of Required Tests



Notation Key

B8e = test accurately predicts student achievement as a selection criterion (1=little importance, 5= great importance)

D1 = factors perceived by teacher as influencing required test results (1=little/no influence, 5=great influence)

D1b = student motivation

D1C = student IQ

D1h = students' true math/reading ability

D48 =reaction to discontinuing required testing (1=negative, 5=positive)

D3 = time cost versus benefit of required tests

A7RA= % of time spent in whole class instruction

TABLE 12

Direct and Indirect Effects on Classroom Context and Attitudes toward Usefulness of Required Tests by Hypothesized Causes

Causes	Direct Effects	Total Indirect Effects	Total Effects of Hypothesized Causes
Whole Class Instruction			
School SES Level	-.279		-.279
Attitudes toward Usefulness of Required Tests			
Whole Class Instruction	-.181	-.031	-.212
Test Quality Concerns	-.733	-.008	- 741

attitudes toward usefulness of required tests included the teachers' opinion of how positive they would feel if required tests were discontinued and how positive they felt about time spent on required tests (too much - too little).

The estimated structural equation parameters are written above each causal path with their t-value in parentheses. Because of the nature of the scaling of the latent variables, negative coefficients appear, but they are all in the predicted direction.

Each path was found to be significant or nearly significant at the .05 level, with the overall model "fit" quite good, as indicated by a chi-square goodness of fit of X^2=14.46, df=22, p=.8845. This represents a significant improvement over the measurement model, x^2=31.6, df=26, p=.2067, wherein the structural relations are set to zero, and over the null model X^2=79.60, df=28, p=.0000, where the number of observed variables is equal to the number of latent constraints, and all measurement error variances are set to zero.

DISCUSSION

INTRODUCTION

Teacher use of tests and test results is a topic that has generated much rhetoric and theory attempting to explain observed nonuse and misuse. Little empirical support has been generated regarding the validity of our perceptions of testing in its relationship to teaching and learning. The study reported here is one of a handful that has investigated what teachers

themselves know and believe about tests, and how this attitude affects use of test data in the classroom.

Our findings lead us to three categories of recommendations: about tests; about teachers; and about future test use research. The first concerns the features inherent in standardized, norm-referenced tests which are frequently employed in mandated testing programs. Of primary concern is the content validity of the tests for making statements about an individual student's level of performance. To increase the usefulness of tests, and hence the actual use of test data, test developers need to determine instructional information needs and to attempt to provide that information, primarily diagnostic/prescriptive, in a meaningful test score or set of scores.

On the other hand, our second category of findings suggests the need for a complementary effort on the part of practitioners; that is, teacher training, both preservice and in-service, must begin to address the now evident inadequacy of teachers' skills in educational testing. Standardized tests, and the external mandate they often serve, cannot adequately serve two audiences (e.g., teachers and funding agencies) with disparate evaluation needs. Comparative standings and data that can be meaningfully aggregated to larger instructional units will remain at odds with teacher need for instructional prescription and diagnosis, referenced to particular classroom activities and content and provided at the individual level. Accordingly, teachers in search of the latter information will need to be able to understand and use information gathered for the former. Test theory training and experience in testing and test interpretation will help increase

the benefits derived from carefully designed standardized testing for the classroom teacher.

In addition to training in the more traditional testing theory, our data suggest a subcategory of recommendations for improving teacher training in testing. Teachers rely heavily upon their own tests, upon curriculum-embedded tests, and most heavily upon informal assessment techniques (observation, interaction). While error costs from these information sources may be low, improving teachers' skills in the development of reliable, sensitive (valid) techniques and instruments is certainly an appropriate recommendation for training in testing/assessment.

Our third set of recommendations addresses future research and study in test use by teachers. Two particular notions emerge from the data reported here. The first is that we must expand our definition of and attention to "tests" and "test use." Clearly, teachers use a variety of information sources beyond the domain of paper and pencil assessment; investigators of classroom processes and teacher decision making have long recognized this fact. However, we must begin to carry some of that understanding into our work in testing, if we are to truly increase the benefits end of cost/benefits ratio associated with testing. The second, related notion in this set of recommendations pertains to the instructional options which are available to teachers and which affect the kind and the frequency of performance information sought.

The study data and specific conclusions are discussed in the following sections.

DISTANCE FROM INFORMATION SOURCE AND THE NEED FOR INFORMATION

Our survey data verify two commonly held beliefs about teachers and testing: required, formal achievement testing is a regular feature of schooling, and test data from externally imposed testing programs are neither useful nor used for instructional purposes. Almost all schools in our sample reported required achievement tests for accountability, evaluation, or "survey" purposes. Teachers themselves reported giving many tests. In contrast to state and local agencies who regularly require norm-referenced standardized achievement tests, teachers reported greater use of program-embedded or teacher-made tests.

On the other hand, our results suggest that although teachers find little use for required, standardized test results, such results <u>are</u> used for two specific purposes: to communicate about students to staff or parents, and to group students by achievement/ability levels at the beginning of the school year. Our findings further suggest that teachers view formal test data as less valid for instructional purposes (in terms of timeliness, instructional relevance, norm group referenced, etc.) than data derived from more intimate, contextualized sources of information such as teacher-made tests, interactions with and observations of students, student work samples, etc.

That the primary instructional use of results from formal, mandated tests is confined to grouping students at the beginning of the school year suggests to us that teachers view test data as supplemental; to be used when there is little else available to indicate students' general

performance levels. By the same token, for other staff members, e.g., specialists, counselors, principals, and for parents who do not have an opportunity to observe or interact with students directly, test results serve an important informational function. Further, for program planning and evaluation, district or state allocation of funding, and for decision-makers who are even more distal from students, information gathered through informal idiosyncratic teacher methods, e.g., oral quizzes, work samples, and observations, is both difficult to reliably document and to legitimately aggregate. Thus, the reliance upon and importance of results from standardized testing becomes even greater.

If we categorize all information-gathering techniques teachers use, formal testing, oral quizzes, and work samples all look at students' products, whereas observation and interaction look at the learning process. Our survey, as well as available literature, suggest that for making instructional decisions, teachers need "diagnostic" information. For this, they are concerned with the "process" more than the product of student learning. However, process information gathered through observation and interaction suffers from the measurement problems of objectivity and difficulty in documentation. Until these problems are alleviated, formal testing will continue to serve an important function in education.

Other consumers of assessment data operate in the realm of summative decisions emphasizing learning product, more than process, information. Further, this category of test users needs aggregate data for making decisions about instructional units larger than individual students and teachers, e.g., programs, schools, districts, grade levels. The

distinction between classroom proximal and classroom distal decisions and decision-makers appears to be important, explaining the difference in perceived utility of various sources and types of "test" or assessment information.

QUALITY OF INFORMATION AND RISKS ASSOCIATED WITH WRONG DECISIONS

The technical quality of teacher-made and curriculum-embedded tests has long been questioned. Our study, however, finds that teachers use results from these tests extensively for making instructional decisions. As discussed in earlier sections on test features and teacher attitudes, teachers believe standardized test results do not adequately reflect classroom instruction and are therefore of limited value. Also, teacher-made and curriculum-embedded tests are administered at the teacher's discretion, when information of particular kinds is needed at particular decision points.

Beyond the explanation of quality, i.e., content validity and timeliness, we suggest that the risks associated with test-based decisions are a factor influencing use. Decisions made on the basis of test results vary in their reversibility. For example, formal decisions include those to select people for jobs or college entrance; to certify achievement for grade-to-grade promotion or high school graduation. Such decisions have a greater impact on individuals and are more difficult to reverse than the less formal decisions. Less formal and easier to reverse decisions include those within the classroom instructional settings, such as unit to unit advancement or grouping of students and assignment of material. These latter teacher decisions are more immediately verifiable and correctable. That is, effects of a

wrong decision can be corrected more readily. Because these classroom decisions are more easily correctable, using information of questionable quality (reliability) is not as serious a problem as it is for making decisions of more permanent consequence.

TESTING AND INSTRUCTIONAL CONTINGENCIES

Within the classroom, information from testing can be used for making decisions about grouping/regrouping, placing students on a continuum, selecting teaching methods and material, and so on. In this sense, level of use of test results is not only influenced by teachers' training and attitude toward testing and instruction, but also by the frequency with which various kinds of decisions arise. These decisions, in turn, are related to the alternative teaching strategies available to the teacher (Shavelson et al., 1977).

However, alternative teaching methods are available to teachers only when teachers perceive them as available. For example, the mere presence of an aide does not mean the teacher views and consequently uses the aide as an instructional option. Our study obtained information on facilities and personnel that constitute potential resources for increasing the range and kind of instructional options. Future studies should verify this definition of range of resources in terms of classroom realities. That is, what does the teacher perceive his/her options to be?

TEACHER KNOWLEDGE AND TRAINING

Inadequacies in teachers' knowledge of and training in test construction and interpretation of results have often been cited as a major reason for teachers' low level of use of test data. Our study agrees with this hypothesis, but finds an interesting distinction between the effects of self-reported training and measurable knowledge in testing. Teachers who had more training in tests and measurement were, in fact, more likely to make use of results from required tests. However, those teachers actually able to demonstrate their knowledge (interpreting percentile and grade equivalency data) were more reserved in their reliance upon such test data.

The literature on extant teacher certification and district in-service programs explains much of the lack of teacher knowledge of and attitude toward tests and testing; indeed, review of this literature makes it difficult to fault teachers for lack of knowledge or for misconceptions about testing. For example, Woellner (1979) provides a national picture of teacher preservice training and certification. On a state by state basis, the certification requirements demand virtually nothing of teachers in terms of formal course work in testing. Professionals providing special services, e.g., the school psychologist or district psychometrist, do have some minimal requirements for course work in individual and group assessment. For some administrative/specialist credentials, an additional year of course work may be required, but rarely does this include formal testing course work.

In the vast majority of the states, there are teacher requirements for course work in teaching methods based on the social or behavioral sciences or general psychology. These general courses are likely to provide little concrete information on tests and testing. Even the descriptions for specialist course work for teachers have few, if any, formal requirements for courses in testing. Overt references to such concepts as testing for diagnostic and prescriptive purposes are non-existent. A few states (e.g., Indiana, Oklahoma, Tennessee) mention, as part of professional course work, training in "evaluation of learning" or "measurement and evaluation." However, since the total number of semester hours required for all professional training is quite small, the amount of time spent specifically on tests and testing is likely to be very limited.

In the field of in-service training, recent emphases make no direct statement of need to provide teachers with training in testing. Rather, such work (Adam, 1975; Johnston, 1971; and Harris, 1980) generally discusses how to design, conduct, and assess in-service operations; it does not prescribe specific in-service content, such as components in tests and testing.

The literature on test use provides different directions that might be taken to improve testing practices and use. For example, Quinto (1977) suggests that teachers should have a deciding voice in specifying the purpose of giving a particular test and how the results should be interpreted and used. Baker (1974) proposes the use of domain-referenced testing. She suggests that behavioral objectives by themselves do not provide sufficient cues to the teacher for altering instruction to improve learning.

Domain-referenced testing is presented as a viable alternative to the vague or trivial test items that often derive from unspecified objectives. Content rules of domain specifications, bound as they are to instructional content, are viewed as a departure from the more common "objectives-based" evaluations.

In addition, Baker (1980), among others, calls for more sweeping revisions in current test selection and development practices to improve test and instruction linkages. She proposes four criteria for educational tests: that they be economical; closely related to instruction; have public accessibility; and offer significant experiences for students.

Most authors on the subject have recognized the need to improve teacher knowledge of tests and testing, to increase their involvement in the testing endeavor, and to facilitate their use of test results; that is, to make teachers more informed consumers of tests and test data. Hastings et al (1961) describe a study in which the attitudes and perceptions of test users can be improved as a result of training. Ebel (1967) has called for in-service training workshops to improve teacher competence in tests and testing. Unfortunately, to date, there has been little actual follow-up to such immediate practical suggestions.

Such training is crucial to teacher understanding and use of test data for instructional decision making. Whether one is discussing standardized tests, local tests, teacher-made or curriculum-embedded achievement tests, or alternative approaches to student assessment, teachers' knowledge and understanding of the particular kind of measure to be administrated, their commitment to use of that measure, and the availability

of linkages between the measure and instructional improvement will be critical (Goslin, 1967). It does not appear that teacher training institutions are providing such training at the preservice level; nor does such training appear to be systematically provided by local districts at the level of in-service. The need for and value of increased teacher knowledge in tests and testing suggests we should reconsider present teacher training priorities and mobilize support and allocate resources toward two goals: 1) making teachers more informed consumers of tests and test data; and 2) strengthening the relationship between teaching, testing, and learning.

REFERENCES

Adam, E. (Ed.) In-service education and teachers' centers. New York: Pergamon Press, 1975.

Ahmann, J. S. National achievement profiles in ten learning areas. Educational Studies, 1979, 9(4), 351-364.

Airasian, P. W., Kellaghan, T., Madaus, G. F., & Pedulla, J. J. Proportion and direction of teacher rating changes of pupils' progress attributable to standardized test information. Journal of Educational Psychology, 1977, 69(6), 702-709.

Baker, E. L. Beyond objectives: Domain-referenced test for evaluation and instructional improvement. Educational Technology, 1974, 14(6), 10-16.

Baker, E. L. The evaluation of the early childhood education program: Volume I. Los Angeles, California: Center for the Study of Evaluation, 1975.

Baker, E. L. Achievement testing in urban schools: New numbers. St. Louis, Missouri: CEMREL Monograph on Urban Education, in press, 1980.

Bauerfeind, R. H. Building a school testing program. Boston: Houghton Mifflin, 1963.

Boyd, J., McKenna, B. H., Stake, R. E., & Yachinsky, J. A study of testing practices in the Royal Oak (Michigan) public schools. Royal Oak, Michigan: Royal Oak City School District, 1975. (ERIC Document Reproduction Service No. ED 117 161)

Broekhoff, M. Criterion-referenced evaluation as an alternative. English Journal, 1978, 67(7), 32-37.

Christiansen, T., & Livermore, G. A comparison of Anglo American and Spanish-American Children on the WISC. Journal of Social Psychology, 1970, 81(1), 9-14.

Cramer, S., & Slakter, M. A scale to assess attitudes toward aptitude testing. Measurement and Evaluation in Guidance, 1968, 1(2).

Crew, J. L., & Whitney, E. N. Criterion-referenced testing: Usages in some member systems of the Council of great City Schools. Journal of Negro Education, 1978, 47(2), 159-167.

Ebel, R. L. Improving the competence of teachers in educational measurement. In J. Flynn and H. Garber (Eds.), Assessing behavior: Readings in educational and psychological measurement. Reading, Mass.: Addison-Wesley, 1967, 171-182.

Gardner, W. E. A model for creating a more hospitable achievement environment for Black elementary students. Negro Educational Review, 1977, 28(3), 229-236.

Gardner, W. E., & Gardner, D. J. Achievement test construction and selection for Black children. Journal of Non-White Concerns in Personnel and Guidance, 1978, 6(3), 148-156.

Glass, G. V When educators set standards. In E. L. Baker and E. S. Quellmalz (Eds.), Educational Testing and Evaluation: Design, Analysis, and Policy. Beverly Hills, California: Sage Publications, Inc., 1980, 185-193.

Gorrow, F. F. Better classroom testing. San Francisco: Chandler Publishing Co., 1966.

Gorth, W. P., & Perkins, M. R. A study of minimum competency testing programs: A final report. Washington, D. C.: Office of testing, Assessment and Evaluation, National Institute of Education, 1979.

Goslin, D. A. The search for ability: Standardized testing in social perspective. New York: Russell Sage Foundation, 1963.

Goslin, D. A. The use of standardized tests in elementary schools. New York: Russell Sage Foundation, 1965.

Goslin, D. A. Teachers and testing. New York: The Russell Sage Foundation, 1967.

Green, D. R. Racial and ethnic bias in achievement tests and what to do about it. Monterey, California: CTB/McGraw-Hill, 1973.

Harris, B. M. Improving staff performance through in-service education. Boston: Allyn and Bacon, 1980.

Hastings, J. T., Runkel, P. J., & Damrin, E. E. Effects on use of tests by teachers trained in a summer institute. Urbana: University of Illinois, Bureau of Educational Research, 1961.

Hilliard, A. G. Cultural variation and living assessment in the service of instruction. Paper in preparation based on the March 1979 meeting of the Study Panel on Assessment, sponsored by the MIT Division for Study and Research in Education.

Hoepfner, R., Conniff, W. A., Jr., Hufano, L., Bastone, J., Ogilvie, V. N., Hunter, R., & Johnson, B. L. CSE school test evaluations. Three volumes. Los Angeles, California: University of California. Center for the Study of Evaluation, 1977. (ERIC Document Reproduction Service No. ED 113 382 384)

Houts, P. L. Standardized testing in America II. The National Elementary Principal, 1975, 54(6), 2-3.

Howe, H., II. Tests and schooling. Washington, D. C.: Paper presented at the National Conference on Achievement Testing, Washington, D. C., 1978. (ERIC Document Reproduction Service No. ED 152 852) (a)

Howe, H., II. Tests and Schooling. Two papers presented at the National Conference on Achievement Testing, Washington, D. C., 1978. (b)

Johnston, D. J. Teachers' in-service education. New York: Pergamon Press, 1971.

Jöreskog, K. G., & Sörbom, D. LISREL user's guide, version IV, release 2. Uppsala, Sweden: University of Uppsala, 1978.

Kahn, A. Remarks to the National Conference on Achievement Testing and Basic Skills, Washington, D. C., March, 1978.

Kauffman, J. D. State assessment program: Current status and a look ahead. Paper presented at the annual meeting of the National Assessment of Educational Progress, Denver, Colorado, 1979.

Kirkland, M. The effects of tests on students and schools. Review of Education Research, 1971, 41(4), 303-350.

Klein, S. P. Evaluating tests in terms of the information they provide. Evaluation Comment, 1970, 2(2), 1-6.

Leiter, K. C. W. Teachers' use of background knowledge to interpret test scores. Sociology of Education, 1976, 49(1), 59-65.

Lortie, D. C. School-teacher: A sociological study. Chicago: The University of Chicago Press, 1975.

Mayeske, G. W. On the explanation of racial-ethnic group differences in achievement test scores. Unpublished manuscript, 1971.

Mercer, J. R. Pluralistic diagnosis in the evaluation of Black and Chicano Children: A procedure for taking sociocultural variables into account in clinical assessment. Paper presented at the meeting of the American Psychological Association, Washington, D. C., September, 1971.

Mercer, J. R. IQ: The lethal label. Psychology Today, 1972, 6(4), 44; 46-47, 95-96.

Mercer, J. R. Labeling the mentally retarded. Berkeley, California: University of California, 1973.

Petrosko, J. M. The quality of standarized high school mathemetics tests. Journal for Research in Mathematics Education, 1978, 9(2), 137-148.

Problems and needed improvements in evaluating Office of Education programs. Report to the Congress by the Controller General of the United States. Washington, D. C.: GAD, HRD 76-165, September, 1977.

Quinto, F. Teacher-made tests--an alternative to standardized tests. Today's Education, 1977, 66(2), 52-53.

Shavelson, R. J., Atwood, N. K., & Borko, H. Experiments on some factors contributing to teachers' pedagogical decisions. Cambridge Journal of Education, 1977, 6, 295-307.

Skinner, B. F. The technology of teaching. New York: Appleton, 1968.

Stetz, F. P., & Beck, M. D. (Psych. Corp., N. Y.) Comments from the classroom: Teachers' and students' opinions of achievement tests. Paper presented at the annual meeting of NCME, San Francisco, 1979.

Walker, C. B. et al. CSE criterion-referenced test handbook. Los Angeles, California: Center for the Study of Evaluation, 1979.

Woellner, E. H. Requirements for certification for elementary schools, secondary schools and junior colleges. (44th ed.) Chicago: The University of Chicago Press, 1979.

Yeh, J. P. Test use in schools. Los Angeles, California: Center for the Study of Evaluation, 1978.